Mathematics for High Dimensional Data

36-747: A Statistical Viewpoint

1 Basic course information

Instructor:	Yuting Wei; `ytwei AT cmu DOT edu`
Lecture:	MW 10:40-12pm; Zoom ID: 807 208 2710 (pwd: on *Canvas*) Feb 1 – Mar 17
Office hours:	by appointment
Announcement & materials:	will be posted on *Canvas*

Synopsis: This is a Ph.D. mini class series in statistics. It contains two separate parts with course number 36-747 and 36-748 respectively. We will cover a selection of methods in modern statistics and optimization that are suitable for large-scale problems arising in data science and machine learning applications. The main goal is to provide motivated students with mathematical tools for working on related research problems.

Prerequisites: This course is intended for Ph.D. students with strong mathematical background. There are no formal prerequisites for these minis, but students are expected to have completed at least one intermediate statistics course (like 10-705 at CMU), and preferably an advanced statistics course and one course on optimization. Students are suggested to be familiar with

1. basics for statistics estimation (loss function, least square estimator, efficiency)
2. basic concentration inequalities (such as Markov, Chebyshev, Hoeffding)
3. basics for optimization (convexity, convergence, line search, gradient methods)

2 Course descriptions

Descriptions: In this mini, we will cover the mathematical foundations of several fundamental learning and inference problems. The attention will be concentrated on designing and understanding statistical procedures that are provably efficient in theory and practice.

Tentative list of topics (some might be assigned as reading materials):

- Spectral methods
- Matrix concentration inequalities
- Compressed sensing and sparse recovery

- Convex Gaussian minimax theorem
- Phase transition and convex geometry
- Low-rank matrix recovery
- Robust principal component analysis

Textbook: There is no textbook for this class. Here are some references and additional papers will be uploaded as the class proceeds.

References:

- *High-dimensional statistics: A non-asymptotic viewpoint*, Cambridge University Press, Martin Wainwright, 2019.
- *High-dimensional probability: An introduction with applications in data science*, Cambridge University Press, Roman Vershynin, 2018.
- *Spectral Methods for Data Science: A Statistical Perspective*, Yuxin Chen, Yuejie Chi, Jianqing Fan, Cong Ma, 2020.
- *An Introduction to Matrix Concentration Inequalities*, Foundations and Trends in Machine Learning, Joel Tropp, 2015.
- *High-dimensional data analysis with sparse models: Theory, algorithms, and applications*, John Wright, Yi Ma, Allen Yang, 2018.
- *Convex optimization*, Cambridge University Press, Stephen Boyd, and Lieven Vandenberghe, 2004.

3 Graded components

There will be one homework (30%) whose questions will be progressively released. Homework for Mini 1 is due on Feb 21. No late homework will be accepted. Please use Latex to typeset your homework and email to me. It is okay to collaborate on the homework but you must hand in your own copy.

Another component of the grades will be a course project (10% for proposal, 60% for final report) for *each* of the minis. This project can either be a literature review or include original research:

- Literature review. We will provide a list of related papers that are not covered in the lectures and you are also free to choose any paper that is related to the course materials. The literature review should involve in-depth summaries and exposition of the chosen paper.
- Original research. It can be either theoretical or experimental, with the approval from the instructor. If you choose this option, you can do it either individually or in groups of two.

Three timestamps for the course project:

- Proposal (Feb 28). Submit a paragraph stating the papers you plan to survey or the research problems that you are interested in. Describe why they are important or interesting, and provide some appropriate references. If you elect to do original research, you are encouraged to connect this project with your current research (but is still related to our course content). Please do not propose an overly ambitious project. You will receive feedback from the instructor.
- In-class presentation (if applicable).
- A written report (Mar 21). You are expected to submit a final project report, up to 5 pages long (not including references), summarizing your findings/contributions.